(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 September 2001 (20.09.2001)

PCT

(10) International Publication Number
WO 01/69911 A2

(51) International Patent Classification[7]: H04N

(21) International Application Number: PCT/US01/07320

(22) International Filing Date: 7 March 2001 (07.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/187,699	7 March 2000 (07.03.2000)	US
60/188,264	10 March 2000 (10.03.2000)	US
09/549,797	14 April 2000 (14.04.2000)	US

(71) Applicant *(for all designated States except US)*: **RELATIVE MOTION TECHNOLOGIES, INC.** [US/US]; 1540 Oak Creek Drive, Suite 203, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANGUAY, Donald, O.** [US/US]; P.O. Box 11269, Stanford, CA 94309 (US). **WILBURN, Bennett, S.** [US/US]; 2250 Latham Street, #34, Mountain View, CA 94040 (US).

(74) Agent: **OLYNICK, Mary, R.**; Beyer Weaver & Thomas, LLP, P.O. Box 778, Berkeley, CA 94704-0778 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

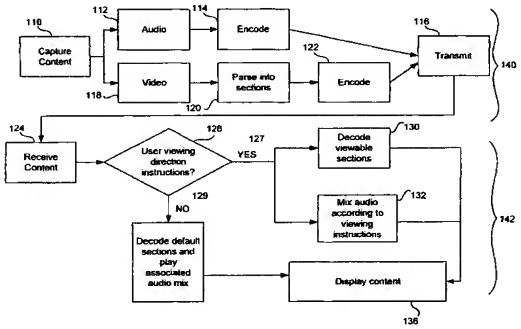
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: INTERACTIVE MULTIMEDIA TRANSMISSION SYSTEM



(57) Abstract: Interactive multimedia system for allowing a user to select a view within a larger visual image. The system includes at least one subscription assembly and at least one transmission assembly. One embodiment of a subscription assembly includes a receiving device, a decoding device, and a rendering device. The receiving device receives encoded video content. The visual content includes animation scripts or a sequence of images. The system parses the image sequence or video content into video streams representing sections of the images. The system individually compresses and formats each of the video streams. The decoding device coupled to the receiving device extracts the individually compressed video streams. The rendering device coupled to the decoding device selectively decompresses the individually compressed video streams and merges the resulting image sections into seamless viewable video. The rendering device includes a user input device for receiving user view selection instructions such that the rendering device renders the animation content and selectively decompresses individually compressed video streams to construct a view based on the user's view selection instructions. Audio channels are also mixed according to user view selection. One version of a transmission assembly for the system includes a content capturing assembly and an encoding device. The content capturing assembly captures video and audio content. The video content includes a sequence of images. The content capturing assembly can include an anamorphic lens to provide a high aspect ratio view without significant warping of the captured video. The encoding device parses the images into video streams representing sections of the images and compresses and formats the video streams for transmission. The transmission is a multimedia stream of audio, video, and graphics content.

WO 01/69911 A2

INTERACTIVE MULTIMEDIA TRANSMISSION SYSTEM

Copyright Notice

Background of the Invention

The present invention relates generally to interactive multimedia transmission systems and, more particularly, to apparatus and methods for allowing a user to interact with transmitted digital content. One such advanced interactive multimedia transmission system is available from Relative Motion Technologies, 791 Tremont St., Boston, MA.

Traditionally, a viewer watching some kind of recorded visual event does so on apparatus having a display screen generally on the order of four units by three units. These numbers dictate an aspect ratio of approximately 1.3 and are commonly seen in things such as traditional motion picture television screens. The physics involved in early motion picture and television production led to the empirical adoption of screens having approximately this aspect ratio.

The aspect ratio of most visual presentation screens, as previously discussed, in combination with the resolution of available cameras which generate the image later watched by viewers, has mandated most television and motion picture format as we know today. In general, a director selects views from within his span of vision for presentation to the audience. One example, for instance as shown at FIG. 1, is a football game. When a viewer watches a televised football game, what he sees are a number of audiovisual clips, each of a few seconds duration which, in the director's opinion, capture the most important moments of the game. In the example shown at FIG. 1, the field of play is shown to be a relatively elongated rectangle, 1, encompassing the field of play having a number of players thereon, generally 5. In the exemplar shown here, the quarterback, 3, is preparing to throw a pass to a receiver,

here 7. The television director has chosen the view of the receiver waiting to receive the ball as the most important view, and that which is being transmitted to the viewing public. This view is shown at 9.

It is of course quite possible that one or more viewers of the game might, instead of watching the receiver, prefer to watch the action where the quarterback is attempting to get the pass off. This of course is not possible. What each viewer watches is mandated by the director without any input from the viewer. Nor is this state of affairs limited solely to televised sporting events. Indeed, most fields of human endeavor are enacted on broad planes of considerable horizontal extent, and we as the viewing public peer at some small fraction of the overall action as though through a keyhole selected for us by another individual whose focus of interest may be entirely different from our own.

Supposing however, that a truly interactive multimedia system were available which could transmit video content to a user and allow the user to interact with the content. In this sense, user interaction could include scrolling or panning across the entire field of play in a football game, or from one side of an opera stage to another. It could include the ability to tilt the view up or down, to be able to see what was happening at the top or bottom of the field of view. Finally, such a system might enable a viewer to zoom in on some feature or individual which he found to be of particular interest.

Heretofore a system that could transmit interactive video content with more than simple detail and with sufficiently high resolution within a practical amount of time would use a considerably larger amount of bandwidth relative to most other types of content transmission systems. In other words, a system that could transmit digital video content uses a relatively large amount of bandwidth. Furthermore, certain systems that transmit high-resolution, interactive digital video content, e.g., spherical immersion systems, generally use even more bandwidth. In light of limited bandwidth, many interactive video transmission systems compress content. As a consequence, a subscription system that is part of such a transmission system must decompress the content for local display or replay.

One such system is described in U.S. Patent No. 5,867,205, issued to McLaren, and incorporated herein by reference. McLaren describes a system for scrolling in a still picture. Another such system is described in U.S. Patent No.

5,990,941, issued to Jackson et al., and incorporated herein by reference. Jackson et al. describe a system for transmitting digital content representing at least one 180 degree field-of-view (FOV). However, transmitting digital content representing at least one 180 degree FOV within a given amount of time uses a relatively large amount of bandwidth.

In addition, many high-resolution interactive video systems use a relatively large amount of computer processing power. Such systems can require processing power to encode and compress video content prior to transmission. Similarly, such systems can require processing power to receive, decode, decompress, and render video content after transmission. The receiving end of a video transmission system is often a device with limited processing power such as a personal computer or a set top box.

Most video systems utilize audio tracks to enhance the visual performance. Where an interactive video system contemplates the incorporation of such audio tracks, provision must be made for the user's changing view. Using the football game exemplar previously presented, the sounds coming from one end of the field might be quite different from the other end. This would certainly be true in the case of a televised play, where the relative volume of the speaker's parts is going to vary on the view. In other words, if a view were viewing the extreme right side of a stage, the sound volume of an actor speaking on the extreme left side of the stage should be lower than if the viewer were focused on him.

What is needed is an interactive multimedia transmission system that enables a viewer to select a view from within a transmitted or recorded video or audiovisual image.

What is also needed is an interactive video system that enables a viewer to perform at least one of scroll, pan, tilt and zoom to enable her to select and focus on the specific aspect of the video presentation in which she is most interested.

What is moreover needed is an interactive multimedia transmission system that is fast and inexpensive while still allowing a user to meaningfully interact with presented content.

What is also needed is an interactive multimedia transmission system that uses less bandwidth than a system that transmits at least one 180 degree FOV.

What is still further needed is a system that balances between the desirability of

high-resolution, interactive content and the associated requirements of a relatively large amount of bandwidth and/or of a relatively large amount of computer processing power.

What is also needed is an interactive video system which incorporates sound tracks intelligently, compensating sound volumes from varying parts of the scene according to the viewer's selected view.

Finally, what is needed is an interactive video system that is usable in a broad range of distribution media types, including but not limited to live broadcast video, recorded broadcast video, and video productions recorded on media. These media include, but are again not limited to videocassettes, compact disks, CD-ROMs, DVD discs, LaserDisks™, and sundry other recording media and memory retention devices, both permanent and erasable.

Summary of the Invention

The present invention teaches an interactive multimedia transmission system that enables, for the first time, a viewer to select a view from within a transmitted or recorded video, multimedia, or audiovisual image. While the principles enumerated herein teach a number of specific embodiments, each of them is capable of at least one of scrolling, panning, tilting and zooming to enable a viewer to select and focus on that specific aspect of the video presentation in which the viewer is most interested.

The interactive multimedia transmission system taught herein is capable of fast transmission times and inexpensive operation, while still allowing a user to meaningfully interact with presented content.

The interactive multimedia transmission system taught herein uses less bandwidth than prior systems that transmit at least one 180 degree FOV.

The interactive multimedia transmission system taught herein balances between the desirability of high-resolution, interactive content and the associated requirements of a relatively large amount of bandwidth and/or of a relatively large amount of computer processing power.

The interactive multimedia transmission system taught herein also incorporates sound tracks intelligently, compensating sound volumes from varying parts of the scene according to the viewer's selected view.

Finally, the interactive multimedia transmission system taught herein is usable in a broad range of media types, including but specifically not limited to live broadcast video, recorded broadcast video, and video or multimedia productions recorded on media. These media include, but are again not limited to videocassettes, compact disks, CD-ROMs, DVD discs, LaserDisks™, and sundry other recording media and memory retention devices, both permanent and erasable.

According to one aspect of the invention, there is provided a method for allowing a user to select a view from within a transmitted video image, wherein the interactive multimedia system includes at least one subscriber system. The method includes the steps of providing encoded video content, decoding individually compressed video streams, receiving view selection instructions from a user, selectively decompressing individually compressed video streams, and rendering seamless viewable video.

The method provides encoded video content to the subscriber system. The video

content includes a sequence of images. Prior to transmission, the images are parsed into video streams representing sections of the images. Each of the video streams is individually compressed and formatted. On receipt of the video streams, the subscriber system decodes the individually compressed video streams and receives view selection instructions from a user. Responsive to instructions received from the viewer, the subscriber system selectively decompresses individually compressed video streams that represent image sections within the user selected view. The system then renders seamless viewable video by rendering and merging the resulting image sections. As used herein, the terms "transmission" and "receipt" specifically include the broad general fields of broadcast images and recorded images.

According to one aspect of the present invention, the video content includes images with normal aspect ratios. Where such normal aspect ratio video is provided with high resolution images, the principles of the present invention enable the interactive multimedia experience previously discussed including, but not limited to, zooming, panning, and tilting.

In another aspect of the present invention, the video content includes images with a high aspect ratio. For present purposes, an aspect ratio of an image is the ratio of the image's width to the image's height. The aspect ratio is preferably greater than 1.7 and most preferably between 3 and 4.

The system taught by the present invention includes at least one transmission assembly for use in preparing the interactive video, and at least one subscription assembly for receiving the interactive video signal. One embodiment of a subscription assembly includes a receiving device, a decoding device, and a rendering device.

One version of a transmission assembly for the system includes a content capturing assembly and an encoding device. The content capturing assembly captures video content. The video content includes a sequence of images. The content capturing assembly can include a camera in operative combination with substantially any lens system known in the art, including but specifically not limited to: spherical lenses, anamorphic lenses, wide-angle lenses, and other lenses known to those having ordinary skill in the art. While according to one embodiment, the camera implemented is a high-resolution DTV camera, substantially any camera capable of capturing a video image may be utilized. Alternative cameras include, but are specifically not limited to cameras having a high aspect ratio sensor, and a 1920x1080i interlaced

HDTV camera

The camera records video content and passes the recorded content in a digital format to the encoding device. The encoding device parses the images into video streams or bands representing sections of the images and compresses and formats the video streams for transmission.

The receiving device receives encoded video content. The term "receiving device", as utilized herein, contemplates not only video devices capable of receiving broadcast signals, e.g. a television set connected to an antenna or cable service, but also video devices capable of receiving therein and playing back recorded video images, including, but again specifically not limited to: videocassette playback devices, DVD players, LaserDisk™ players, CD players, set top boxes, video gaming systems and the like.

The video content includes a sequence of images. The receiving device receives the encoded video content that has been previously been parsed into video streams representing sections of the images. Furthermore, each of the video streams has been individually compressed and formatted. These video streams may, again, be either broadcast or recorded.

The decoding device is coupled to the receiving device for extracting the individually compressed video streams. The rendering device is coupled to the decoding device for selectively decompressing the individually compressed video streams and for merging the resulting image sections into seamless viewable video. The rendering device includes a user input device for receiving user view selection instructions such that the rendering device selectively decompresses individually compressed video streams to construct images in accordance with the user's view selection instructions.

According to another aspect of the present invention, the interactive video system further includes a first microphone for capturing first audio content, and a second microphone for capturing second audio content. The microphones can be directional microphones. Alternatively, a system according to the present invention could include more than two microphones for capturing audio content. The encoding device encodes the audio content into encoded audio streams and then interleaves the encoded audio streams with the encoded video streams. The rendering device then alters the audio content associated with a video stream based on user view selection

instructions.

One embodiment of a method according to the invention includes the steps of capturing video content, parsing the content into streams, and encoding the streams. The video content includes a sequence of images. The parsing step parses the images into video streams or bands representing sections of the images. The encoding step encodes the video streams for transmission.

Another embodiment of a method conducted in accordance with the teachings of the present invention includes the steps of capturing audio content, and encoding the audio content. The capturing step captures first audio content using a first microphone, and second audio content using a second microphone. Preferably, the microphones are directional microphones. Alternatively, the method can include capturing audio content using more than two microphones. The encoding step encodes the first audio content and the second audio content into encoded audio streams, and interleaves the encoded audio streams with the encoded video streams. Subsequently, the system can adjust the mix of the first audio content and the second audio content based on the user view selection instructions.

According to yet another aspect of the present invention, provision is made for the inclusion of animation scripts within the encoded multimedia stream. Animation scripts are the instructions for creating, or rendering, animation images. Being smaller than the actual animation image sequence, animation scripts generally require much less storage and bandwidth than the animation image sequences, which are essentially video content.

According to yet another aspect of the present invention, the generated multimedia stream may further include message banners, which may be placed, for instance, at the periphery of the visual content. According to this embodiment, when the selected view is panned or tilted beyond the visual content, the messages, for instance advertisements, are brought into the user's view. As a further alternative to this embodiment, messages or advertisements could be directly inserted into the visual content.

The director's viewing parameters, or "director's cut" are the default view presented to the viewer. The viewing parameters include the pan, tilt and zoom of the director's preferred view. In a system with multiple visual streams, the director's view may also include the currently selected visual stream.

These and other advantages of the present invention will become apparent upon reading the following detailed descriptions and studying the various figures of the Drawing.

Brief Description of the Drawing

For more complete understanding of the present invention, reference is made to the accompanying Drawing in the following Detailed Description of the Invention. In the drawing:

FIG. 1 is an example of the "view-within-video" property enabled by the present invention;

FIG. 2 is a block diagram of one embodiment of an interactive video system according to the invention;

FIG. 3 is a flow diagram of the operation of an interactive video system according to the invention;

FIG. 4 is a block diagram of the transmission system of FIG. 2;

FIG. 5 shows one embodiment of the encoding device of FIG. 2;

FIG. 6 is a flow chart for one embodiment of the compressor of the encoding device of FIG. 3;

FIG. 7 is a schematic of a MPEG specific content stream transmitted by the interactive video system of FIG.2;

FIG. 8 is a schematic picture of uniform video banding performed with the system of FIG. 2;

FIG. 9 is a schematic picture of non-uniform video banding performed with the system of FIG. 2;

FIG. 10 is an illustration of the audio mixing performed by the system of FIG. 2;

FIG. 11 illustrates the determination of the viewing direction parameter;

FIG. 12 is a block diagram of a subscription system of FIG. 2;

FIG. 13 is a block diagram of the decoding and rendering assembly of the

system of FIG. 2;

FIG. 14 illustrates the operation of the video decompressor portion of the decoding and rendering assembly of FIG. 13;

FIG. 15 shows one of the displays of FIG. 2 including an orientation inset screen;

FIG. 16 is a schematic illustration of user-controlled panning achieved using the system of FIG. 2;

FIG. 17 is a schematic illustration of user-controlled panning and tilting achieved using the system of FIG. 2.;

FIG. 18 is a schematic illustration of zooming achieved using the system of FIG. 2.; and

FIGS. 19a and 19b are a schematic illustration of peripheral advertising formed using the system of FIG. 2.

Reference numbers refer to the same or equivalent parts of the invention throughout the several figures of the Drawing.

Detailed Description of the Invention

Referring now to FIG. 1, the "view-within-video" aspect of the present invention is explained as follows: According to one embodiment of this aspect, a substantially high aspect ratio image, 1, has been transmitted to a viewer, not shown. It will be appreciated that the high aspect ratio image, 1, is but one embodiment of the present invention. Optionally, the principles enumerated herein may, with equal facility, be implemented on substantially any conceivable aspect ratio image, including but not limited to the previously discussed "4x3" low-aspect ratio image, presuming that the resolution of that image enabled the features and advantages taught herein.

Having continued reference to FIG. 1, the viewer has selected a portion of the video image, 9, for viewing. It will be appreciated that this aspect of the present invention implements an interactive video system which enables the user to interact with the content presented. An overview of one embodiment which enables the features and advantages illustrated in FIG. 1 is shown at FIG. 2. Having reference to that figure, an interactive video system, 20, according to the present invention, includes at least one transmission system, 140, and at least one subscription system, 30. In the exemplar presented herein, a plurality of subscription systems, 30a and 30b, is shown. As will be obvious to those having ordinary skill in the art, the principles of the present invention are specifically applicable to the broadcast or distribution of a far more extensive plurality of subscription systems. The principles of the present invention specifically contemplate such extensive broadcast or distribution.

A user, or viewer, can select a view within presented content by using a user input device 38a to send instructions to the decoding and rendering assembly 34a. Selecting a view can include panning, tilting and zooming, and most preferably includes at least horizontal panning. Visual content is defined herein as video content or scripts, which specify a digital animation sequence. Video content, as used herein, is a sequence of images, obtained, for example, by optically recording a physical event such as a football game or a soap opera. Alternatively, the sequence of images can be generated content, including the rendered output of animation scripts. As a further alternative, a director at the transmission system can specify a default view within the video. Data specifying the default view, or "director's cut" may be transmitted with the presented content, to enable a viewer to select between viewing the director's cut and selecting his own view.

Decoding and rendering assembly 34 may be implemented as a computer, a set top box, a game console, or other means well known to those having ordinary skill in the art for controlling an image on a video display unit. One set top box is available from Scientific Atlanta. By way of illustration but not limitation, an example of a game console is a Sony Playstation 2™. It will be appreciated that these are exemplar devices, and other known devices capable of decoding and rendering in accordance with the principles enumerated herein may, with equal facility, be implemented. User input device 38 may be implemented as a mouse, keyboard, joy stick, game controller, remote control, video camera with computer vision, or other input devices well known to those having ordinary skill in the art. By way of illustration but again not limitation, display device 36 may be implemented as a computer monitor, television screen, HDTV television set, projection display, head-mounted display (HMD), or other visual display device known to those having ordinary skill in the art.

Having reference now to FIG. 3, the operation of system 20 is outlined as follows: Transmission system, 140, obtains content, 110, utilizing content capturing assembly 22. Content capturing assembly may include one or more cameras or microphones, not shown in this view. The audio-visual signal of content, 110, is treated as follows: The audio portion of the signal is split out at 112 and encoded at 114 before being transmitted at 116. The video signal, at 118, is parsed into sections at 120 and encoded at 122. The encoded video signal is then transmitted, along with the encoded audio signal from step 114, at step 116. The encoded audio signal 114, and video signal 122 are transmitted to subscription assembly, 30.

As previously discussed, the principles of the present invention specifically contemplate the implementation thereof on a wide variety of audio, video, and audio-video methodologies. These methods include, but are specifically not limited to, live broadcast, pre-recorded broadcast, netcast, and recorded media. The term "recorded media" in turn specifically contemplates substantially any known methodology for recording audio, video, or audio-visual signals, including but again specifically not limited to magnetic tape, video cassettes, laser disks, compact disks, LaserDisk™, DVD, as well as other magnetic, optical, and electronic storage methodologies well known to those having ordinary skill in the art. The term "netcast" refers to any of several known technologies for transmitting audio, video or audio-visual signals over a network, including Internet. By way of illustration, but not limitation, one such

netcast technology implements a digital broadcast embedded into analog NTSC television signal and is available as Intercast™ from Intel™ Corporation. Alternative netcast technology is available from Webcasts.com™.

Having continued reference to FIG. 3, once the audio-visual signal or media stream is transmitted at 116, it is received, at 124 by a subscription system, 30. At step 126 a determination is made whether to view the "director's cut", or default view within the transmitted video image, or whether the user will provide viewing directions. If, at 128 a decision is made to use the director's cut, the subscription system, 30, decodes the default sections of the transmitted audiovisual signal at 134, and the default content is displayed at 136.

Alternatively, if the viewer decides that she wishes to provide her own viewing directions, a determination is made at 127 and procession of the received encoded audiovisual signal proceeds. At step 130 the desired viewing sections are decoded. At step 132 the encoded audio signals are mixed in accordance with the viewing instructions provided by the user. The decoded audio and video signals are then merged and transmitted to the display unit, and the decoded content displayed, again at step 136. Viewing directions are commands translated from input device 38 in a manner well known to those having ordinary skill in the art.

Having reference now to FIG. 4, transmission assembly 140 of system 20 includes a content capturing assembly 22 including a lens 28 and one or more microphones 29. In the embodiment presented in this figure, there are three microphones, 29a, 29b, and 29c. Content capturing assembly, or camera, 22 utilizing lens 28 captures video content having a field of view. As will be described further below, the illustrated transmission system can capture or provide digital video content that is larger than a user's display. Preferably, the video content includes images with high aspect ratios. The aspect ratio of an image is the ratio of the image's width to the image's height. Standard television has an aspect ratio of approximately 1.3. High-definition television (HDTV) has an aspect ratio of approximately 1.7. One version of a system according to the present invention captures and/or provides images with aspect ratios of preferably greater than 1.7 and most preferably between 3 and 4.

According to the illustrated embodiment, the content capturing assembly includes a lens 28 coupled to a high-resolution video camera 22. The camera assembly 22, 28 preferably provides a high aspect ratio image to achieve a panoramic effect. A

system according to the invention preferably provides high-resolution video with a vertical resolution of 480 pixels or greater and an aspect ratio of greater than 1.7 and most preferably between 3 and 4. The maximum aspect ratio corresponds to a complete 360 degree panorama. During scrolling, a preferred system displays each view with 640x480 (VGA) resolution. An aspect ratio of 4 with 480 vertical pixels implies 1920 horizontal pixels. The present invention contemplates several embodiments for achieving high aspect ratio content.

According to one embodiment, an image sensor, e.g., a CCD or a CMOS imager, captures high aspect ratio video content. A special lens is not necessary. However, the system can use a wide-angle lens to provide a wide angle field-of-view. According to this embodiment, the camera is preferably a progressive scan camera because interlacing can introduce artifacts in the presented video content.

According to another embodiment, the video content includes digitized film recordings. The digitized film recordings can be cropped at the top and bottom to create high aspect ratio video content. The resolution of film is presently higher than the resolution of digital video content provided by current image sensors. Therefore, digitized film recordings can have a vertical resolution of 480 pixels or more, after cropping.

According to yet another embodiment, the system uses a 1920x1080i interlaced DTV camera. Either the camera or the content can be modified to provide high aspect ratio content.

The system can generate 1920x540 interlaced video by cropping a portion of the video. Interlaced video is composed of two fields. Each field contains every other horizontal line of the image with odd numbered lines in one field and even numbered lines in the other field. The video content consists of alternating even and odd fields at 60 fields per second, or 30 frames, consisting of both fields, per second.

Interlacing has advantages. Interlacing allows the display content to be updated more frequently. Second, conventional televisions display NTSC video. NTSC video is interlaced. However, interlaced video can have noticeable artifacts, such as "jaggles". Furthermore, computer monitors preferably use progressive scan video. It is contemplated that the principles enumerated herein may, be implemented on either interlaced or progressive scan video.

Alternatively, the system can store only one field from the 1920x1080i

interlaced camera to produce a 1920x540 image. This system provides 30 frames per second of progressive scan video. Interlacing can be undesirable for the present system because it can complicate the compression and rendering process. Keeping only one field at 30 Hz halves the vertical resolution to create the effect of doubling the horizontal resolution. In other words, the camera decreases the total image size by downsampling the vertical dimension. Thus, the camera halves the height of recorded images.

In a preferred embodiment, the system compensates for the reduction in image height by using an anamorphic lens 28 that focuses a vertically stretched image on the camera sensor. In this embodiment, the image is vertically stretched to compensate for the above-described camera qualities. An anamorphic lens produces different optical magnification along mutually perpendicular axes. In a preferred embodiment, the lens magnifies the vertical dimension of the image by two times, while leaving the horizontal dimension unaltered.

An anamorphic lens provides advantages over a fish-eye lens. An anamorphic lens can induce less optical distortion on the captured images than a fish-eye lens. Further, an anamorphic lens can induce a slight vertical blur to avoid aliasing when vertically downsampling. The final result is viewable high-aspect ratio video content with minimal warping.

A system according to the invention can use an anti-aliasing filter as an alternative to an anamorphic lens. Preferably, the anti-aliasing filter produces a pre-selected blur along the vertical axis.

According to still another alternative embodiment, a system according to the invention can extend the horizontal field-of-view (FOV) to a 360-degree panorama with limited vertical resolution, e.g., 480 pixels. The system can achieve a high aspect ratio image, including a complete panorama, by horizontally merging more than one camera output. One can understand a complete panorama as an equatorial slice of an image sphere.

After the content is captured by capturing assembly 22, the resultant audiovisual signal, or media stream, is transmitted to encoder 24. The encoded media stream is then forwarded to a transmission assembly 26 and then distributed to one or more subscription systems.

Transmission assembly 26 transmits the media stream in a format appropriate to

the receiving subscription system. By way of illustration, but not limitation, transmission assembly 26 can include a television broadcast transmitter, videocassette recorder, CD recorder, and substantially any other broadcast, inter-cast or media recorder known to those having ordinary skill in the art.

Transmission system 140 may further include one or more sources, 27, of generated content. By way of illustration, but again and not limitation, generated content includes but is specifically not limited to the director's viewing parameters, animated figures, designs, messages, advertisements, and other artificially created or previously recorded audio or visual content as an alternative or supplementary content source. After the generation, by generated content source, 27, of an element of generated content, that element or elements are encoded in an encoder, 27a prior to being sent to transmission assembly 26. Encoder 27a may be substantially similar to encoder 24.

Having reference now to FIG. 5 the operation of an exemplar encoder, for instance 24, is discussed. The video and audio signals from content capturing assembly 22 are transmitted to encoder 24. The video signal is first transmitted to a splitter 150. Splitter 150 splits the incoming video stream into a plurality of video streams. The plurality of video streams corresponds to the number of encoded vertically split video bands, as will be hereinafter discussed. Each of the split video streams is then fed into a compressor, 152. In the exemplar presented in FIG. 5, this is a compressor capable of motion compensated video compression. One example of such compression is MPEG2. For the purposes of illustrational succinctness, only one compressor 150 is shown in this view. While the principles of the present invention are applicable to a compressor capable of simultaneous multiple channel video compression, one aspect of the present invention contemplates the utilization of one compressor, 152, for each video stream transmitted from splitter 150. By way of illustration, but not limitation, splitter is capable of splitting an incoming video stream into a plurality of video streams include, but are not limited to, a Microsoft™ Direct Show™ filter.

Following compression of the several video streams by compressor 152, each of these streams is then transmitted to packager 154. Packager 154 may be implemented as a Microsoft™ Direct Show™ filter. Also capable of implementation at this level are the injection of text and director's cut information into the packaged media stream.

This is accomplished by injecting text from a text source 156 directly into packager 154. In similar fashion, the director's viewing parameters, messages and advertising banners and animation scripts, as previously discussed, may also be injected directly into packager 154 from 158, 160 and 162 respectively.

Messages, advertising banners and the like can be joined with video content, as shown in FIG. 19a to provide substantially "circular" content. The results of such circular content are shown at Fig. 19b.

Having reference now to FIG. 6 an exemplar compressor is shown. Suitable compressors include, but are specifically not limited to, computers, including Silicon Graphics™ or Intel™ computers implementing MPEG2 or Intel™ Indeo™ motion compensated video compression schemes. Alternative compressor schemes or methodologies well known to those having ordinary skill in the art may, with equal facility, be implemented. One of the split video streams from splitter 150 is received at converter 42 of compressor 152.

Having continued reference to FIG. 6, a known compressor methodology suitable for implementation as the compressor of the present invention is shown. In this example the compressor methodology is MPEG2. This embodiment is a motion-compensated system. Video received from the content capturing assembly 22 passes concurrently to the motion estimator 56 and to the line scan to block scan converter 42. The motion estimator 56 compares an incoming frame with a previous frame stored in the frame store 62 in order to measure motion, and in order to send the resulting motion vectors to the prediction encoder 60. Motion estimation is performed by searching for the best block-based correlation in a local search window that minimizes a difference metric. Two common methods are the minimum squared error (MSE) method and the minimum absolute difference (MAD) method.

The motion estimator also shifts objects held in the frame store output to estimated positions in a new frame, a predicted frame. The predicted frame is subtracted 44 from the input frame to obtain the frame difference or prediction error. The frame difference is then processed with a combination of DCT and quantization. There is also a local decoder within the encoder that performs a dequantization 58 and an inverse DCT 66. Thus, the local decoder adds the locally decoded prediction error

to the predicted frame to produce the original frame (plus quantizing noise). The resulting frame updates the frame store 62.

Finally, frame difference information and motion vectors are encoded and forwarded to transmit buffer 54. Compression systems often control the rate at which they transmit information. Thus, the system can use a transmit buffer 54 and a rate controller 50 to provide content information at a controlled rate. The rate controller 50 monitors the amount of information in transmit buffer 54 and changes the quantization scale factor appropriately to maintain the amount of information in the transmit buffer between pre-selected limits. The content information is then transmitted from the transmit buffer to the receiver 32 of FIG. 1. The receiver forwards the received content information to the decoding and rendering assembly 34.

According to a preferred embodiment, the encoding device includes a personal computer running a streaming media processing application such as Microsoft's DirectShow™. However, the encoding device can include a personal computer running one of a variety of media processing applications known to those skilled in the art. The system according to the invention can use different compression/decompression schemes, including MPEG, MPEG2, and Intel Indeo™ compression/decompression schemes.

As noted above, a system according to the invention, can use different compression/ decompression schemes. FIG. 7 illustrates an MPEG specific content stream that can be transmitted by the system. The sequence layer 90 contains among other information a variety of video sequences. A video sequence can contain a number of groups of pictures (GOPs), as illustrated in the GOP layer 92. A GOP can contain Intra (I), Predicted (P), and Bi-directional Interpolated (B) frames, as illustrated in the frame layer 94. I frame information can contain video data for an entire frame of video. An I frame is typically placed every 10 to 15 frames. A frame information stream contains a number of macroblock (MB) information streams, as shown in section layer 96. A macroblock information stream in turn can contain MB attribute, motion vector, luminance and color information, as shown in macroblock layer 98. Finally, the luminance information, for example, can contain DCT coefficient information 100.

Independent compression of bands can result in poor image reconstruction at the borders between the bands, in part because motion estimation can fail. Therefore,

reducing the number of borders and providing motion estimation across bands minimizes the problem of poor image construction at the borders between bands.

Conversely, parsing video content into smaller sections or bands reduces the amount of processing that the system performs to create the viewed video. In other words, the system decompresses less total video content to render a selected view when the individual bands are smaller. MPEG blocksize is the minimal practical band size.

Once the content is captured or provided, one embodiment of the system 20 encodes the content, as previously discussed. The system can encode the content by parsing the images that make up the content into sections. The process of parsing images into sections can be termed video banding. A system according to the invention can perform uniform and nonuniform video banding.

FIG. 8 illustrates uniform video banding. The following discussion assumes a video width normalized to 1. In this example, the video width has been split into 8 bands by the encoder, as previously discussed. A is the horizontal width of the fraction of video in one section or band. S is the horizontal width of the fraction of the transmitted video on the screen or display, i.e., the viewed video. D is the horizontal width of the fraction of the transmitted video that is decoded, i.e., the immediately viewable video.

Alternatively, a system according to the present invention can perform non-uniform video banding, as shown in FIG. 12. Sectioning with non-uniform width bands can be useful for rectangular video because the center of the video has a higher probability of being seen than the sides. For example, if the bands are columns, the left most column of pixels is seen only when the user selects the left most view. Taking this consideration into account, one embodiment of a system according to the invention can parse the video content non-uniformly, with the center bands being wider than the edge bands. Furthermore, the system can devote greater compression resources to the center bands.

In the embodiment illustrated in FIG. 1, the content capturing assembly also includes microphones 29a, 29b, 29c for capturing audio content. In a preferred embodiment, the microphones 29 are directionally dependent, i.e., the microphones are more sensitive in the direction in which they are pointed. A system according to the invention can include any number of microphones including no microphones.

However, in a preferred embodiment, the system includes at least two microphones. The system then mixes the audio content based on user view selection instructions. This feature is explained having continued reference to FIG. 2, as well a to FIG. 10.

According to the embodiments illustrated in FIGS. 2 and 10, the system mixes the audio obtained by microphones 29a, 29b, and 29c of FIG. 1 based on user view selection instructions. In other words, if a user selects a view centered in the direction of microphone 29c, the audio signal from microphone 29c is more heavily weighted than the audio signals from microphones 29a and 29b. Such audio mixing further increases the user's ability to affect her viewing experience and further increases a user's ability to interact with presented content.

One embodiment of a system according to the present invention mixes two audio channels. According to this embodiment, the method of audio mixing varies the volume of each channel according to the view selection instructions, while maintaining a minimum volume based on a predetermined ambient constant Ka. In normal operation, Ka ranges from 0 to 1/3. One embodiment of a two-channel audio mixing method uses the following mixing equation: Audio(d) = (1-(2 Ka))[(1-d)L +d*R] + Ka (L+R), where L is the left audio channel signal amplitude, R is the right audio channel signal amplitude, and d is a viewing direction parameter as described below. The first term is a linear interpolation term and the second term is an ambient term.

FIG. 11 illustrates the viewing direction parameter. For rectangular content, viewing direction can be represented by a parameter d, which varies from 0 (left) to 1 (right). The full content width is Wc. The constant screen width is Ws (in units of pixels). The parameter d varies from 0 to 1 over a range of Wc - Ws in content units. In other words, the center pixel of the current screen at time t has a coordinate Xt that varies from Ws/2 to Wc-Ws/2. Knowing Xt, we can find d at that instant in time: dt = (Xt - 1/2Ws)/(Wc-Ws).

In an embodiment of the present invention which implements zooming capability, the "viewed contents" window can have a size smaller than the screen width W_s. In this embodiment, the center pixel of the current view will have an X-coordinate, Xt, outside of the range [Ws/2, Wc-Ws/2]. Where this occurs, we clamp Xt to this range prior to calculating D_t. In this manner, Dt is always in the range [0,1]. The effect of zooming is shown at FIG. 18.

In an alternative embodiment, a system according to the present invention can mix audio from three channels. In this embodiment, the system mixes audio according to the following equations: Audio(d) = Ka(L+M+R) + (1-3 Ka) * [(1-2d)L + 2d*M] for 0<d<1/2 and Audio(d) = Ka(L+M+R) + (1-3 Ka) * [(2-2d)M +(2d-1)R], for 1/2 < d < 1.

As a further alternative, the principles taught herein may mix a larger number of audio channels, by expanding the previously provided equation to include such larger number of channels.

Having continued reference to Fig. 2 as well as to FIG. 12, a subscription system includes a receiving assembly 32, a decoding and rendering device 34, a user input device 38, and a display 36. In a preferred embodiment the system 20 includes a plurality of such subscription systems. A plurality of subscription systems allows multiple simultaneous users. The present invention enables each user to select his or her view independent of all other users. Thus, for example, if the system records a wide-angle view of a theatrical presentation, one user can center her view on a character on the left side of the stage while another user can center his view on a character on the right side of the stage.

This feature is illustrated having continued reference to FIG. 2, wherein the illustrated image includes, from left to right, a dog, a turtle, and a rabbit. A first user has provided view selection instructions via user input device 38a to decoding and rendering device 34a to center the view between the turtle and rabbit as shown on display 36a. A second user has provided view selection instructions via user input device 38b to decoding and rendering device 34b to center the view between the dog and turtle as shown on display 36b. Thus, the view on display 36b is centered to the left of the view on display 36a.

One exemplar decoding and rendering assembly 34 is further illustrated having reference to Fig. 13. Having reference to that figure, the incoming streaming media received from receiver 32 is received at unpackager 160. Unpackager 160 separates the incoming video and audio signals from the received streaming media. Responsive to the current view selected by the viewer, selector 162 and mixer 164 provide the proper signals for viewing and hearing. Selector 162 determines which of the incoming video streams corresponding to the selected bands are to be decompressed and eventually viewed. In similar fashion, mixer 164 mixes the incoming audio

signals to provide a mixed audio signal again responsive to the users selected view, as previously discussed. After the bands appropriate to the selected view have been selected by selector 162, they are decompressed by decompressor 166. The several bands are then aggregated by aggregator 168 into a viewable video stream. The video stream is cropped at cropper 170 in order to remove unviewed video. The resultant viewed video is then transmitted to display 36.

The decompressor 166 is the logical inverse of encoder 24, previously discussed. This decompressor is detailed having reference to Fig. 14. The incoming video stream equated to one of the bands selected for viewing by the viewer is received at variable length decoder 72. Variable length decoder 72 outputs the resultant signal to a dequantize step 74 and to prediction decoder 82. Prediction decoder 82 transmits the discrete cosine transform coefficient to the inverse discrete cosine transform 76. Prediction decoder 82 further transmits a signal to motion predictor 84. The motion predictor 84 shifts the frame store 86 output by the transmitted motion vectors received from the prediction decoder 82. The result is the same predicted frame as was produced in the encoder. The system then adds the decoded frame error (received from the decoder 72, dequantizer 74, and the Inverse DCT 76) to the predicted frame to produce the original frame.

Decompression of a video stream can begin on an I frame. An I frame begins a group of pictures (GOP) consisting of about 12-15 frames. Given the content streams at 30 frames per second, decompression can begin only once every 1/3 to 1/2 second. The bands must be large enough to allow reasonable velocity within the current bands before new bands are started. For rectangular video, a full pan in 2-3 seconds is reasonable. A full pan is defined as a pan from one extreme end of the rectangular video to the other extreme end. Circular video implies a 4-6 second pan. According to a preferred embodiment, for 4:1 video on a 4:3 screen, the system parses the video content into 7-12 bands. Using 7 bands implies that 4 bands are decompressed, resulting in a 43% reduction of work relative to total decompression. Using 12 bands implies that 6 bands are decompressed, resulting in a 50% reduction of work relative to total decompression.

In an alternative embodiment, display 36b can include an orientation inset screen or "thumbnail" 110, as shown in FIG. 10. The inset screen 110 includes a present display location box 112 for indicating the location of the present selected view within

the larger available visual content. In one embodiment, the present display location box 112 becomes smaller if the user chooses to zoom in on a portion of presented content. A preferred embodiment allows zoom capability to provide one to four times magnification. The area of the inset screen 110 outside the location box 112 can be blank or can show the rest of the visual content as illustrated, in order to provide a "macroview".

Limiting user camera control to scrolling control reduces expense and computational complexity while maintaining interactivity, especially when supplemented with audio mixing as described below. According to an alternative embodiment, the system provides panning control. Panning and scrolling are somewhat similar visual presentations. Panning is defined herein as the rotation of a camera about a vertical axis, and collaterally, a panned image is the image resulting from such movement. A rectangular image is scrolled where it is moved horizontally or vertically so as to present a viewer with an apparently moving segment of the rectangular image. Essentially, the display appears to "slide across" the larger content. Panning control can require perspective correction when the captured content exhibits perspective distortion caused by the camera lens. This correction may be performed by means of normal perspective correction methodologies including, but not necessarily limited to, perspective projection equations, image warping, and other methodologies well known to those having ordinary skill in the art.

Thus, according to one embodiment, a user can scroll as illustrated in FIG. 16. The system can provide the user with fine panning or scrolling control such that section 103 has a width as small as one pixel. In other words, in one embodiment, a user can control her view in the lateral or horizontal direction to an accuracy of one pixel width. In an alternative embodiment, a user can both pan and tilt as illustrated in FIG. 17. If the system does not receive user view selection instructions, the system decodes default sections 134 and displays a default view. Alternatively, according to a preferred embodiment, a user can proactively select a view. Where the user's system enables substantially "VCR-like" functionality including play/pause/resume, the user can pause the action to inspect the entire transmitted image via panning, tilting and zooming.

From the preceding discussion, it will be appreciated that the principles of the present invention enable, for the first time, a truly interactive view-within-video

experience without the need for such excessive bandwidth as would preclude the fielding of the invention.

The present invention has been particularly shown and described with respect to certain preferred embodiments of features thereof. However, it should be readily apparent to those of ordinary skill in the art that various changes and modifications in form and detail may be made without departing from the spirit and scope of the invention as set forth in the appended claims. In particular, the principles of the present invention specifically contemplate the incorporation of one or more of the various features and advantages taught herein on a wide variety of compression/decompression methodologies, cameras, A/V media, distribution methods, and software/hardware platforms. Each of these alternatives is specifically contemplated by the principles of the present invention.

CLAIMS

What is claimed is:

1. An interactive video system for allowing a user to select a view, the system comprising

a receiver for receiving encoded video content, the video content including a sequence of images, the video content having been parsed into video streams representing sections of the images, each of the video streams having been individually compressed and formatted;

a decoder coupled to the receiver for extracting the individually compressed video streams; and

a rendering device coupled to the decoder for selectively decompressing the individually compressed video streams and for merging the resulting image sections into seamless viewable video, the rendering device comprising user interface means for receiving user view selection instructions such that the rendering device selectively decompresses individually compressed video streams to construct a view based on the user's view selection instructions.

2. The interactive video system of claim 1, wherein the system further comprises:

a capture device for capturing the video content, the video content including a sequence of images; and

an encoder for parsing the images into video streams representing sections of the images and for compressing and formatting the video streams into encoded video streams.

3. The interactive video system of claim 2, wherein the capture device comprises:

a first microphone for capturing first audio content, and

a second microphone for capturing second audio content,

wherein the encoder encodes the first and second audio content into encoded audio streams and interleaves the encoded audio streams with the encoded video streams.

4. The interactive video system of claim 3, wherein the rendering device adjusts the audio content associated with a video stream based on user view selection instructions.

5. The interactive video system of claim 2, wherein the capture device comprises an anamorphic lens.

6. The interactive video system of claim 2, wherein the encoder formats the video content into MPEG compliant bit streams.

7. The interactive video system of claim 1, wherein the video content includes high aspect ratio video content.

8. The interactive video system of claim 7, wherein the high aspect ratio video content provides a horizontal angular field-of-view greater than 50 degrees.

9. The interactive video system of claim 7, wherein the high aspect ratio video content provides a horizontal angular field of view greater than 100 degrees.

10. The interactive video system of claim 1, wherein the rendering device decompresses only video streams representing image sections that constitute a user selected view.

11. The interactive video system of claim 1, wherein the sections are non-overlapping, vertical columns.

12. The interactive video system of claim 1, wherein the encoded video content is created and transmitted in real-time.

13. The interactive video system of claim 1, wherein the encoded video content is derived from a storage medium.

14. The interactive video system of claim 2, wherein the system further comprises transmitting means for transmitting the compressed and formatted video streams to the
receiver.

15. A method for allowing a user to select a view, wherein the interactive video system includes at least one subscriber system, the method comprising the steps of:

providing encoded video content to the subscriber system, the video content including a sequence of images, the images having been parsed into video streams representing sections of the images, each of the video streams having been individually compressed and formatted;

decoding the individually compressed video streams;

receiving view selection instructions from a user for selecting a view;

decompressing individually compressed video streams that represent sections of the selected view; and

rendering seamless viewable video by rendering and merging the resulting sections of the selected view.

16. The method of claim 15, wherein the method providing step comprises the steps of:

capturing the video content, the video content including a sequence of images;

parsing the images into video streams representing sections of the images; and

encoding the video streams into encoded video streams.

17. The method of claim 16, wherein the step of capturing comprises the steps of capturing first audio content using a first microphone, and

capturing second audio content using a second microphone, and wherein the encoding step comprises the steps of

encoding the first audio content and the second audio content into encoded audio streams, and interleaving the encoded audio streams with the encoded video streams.

18. The method of claim 17, wherein the rendering step includes the step of

adjusting the mix of the first audio content and the second audio content based on the user view selection instructions.

19. The method of claim 16, wherein the capturing step includes capturing the

video content using an anamorphic lens.

20. The method of claim 16, wherein the compressing step compresses the video streams into MPEG compliant streams.

21. The method of claim 15, wherein the video content includes high aspect ratio video content.

22. The method of claim 21, wherein the high aspect ratio video content provides a horizontal angular field of view greater than 50 degrees.

23. The method of claim 15, wherein the decompressing step decompresses only the video streams representing image sections that constitute a user selected view.

24. The method of claim 15, wherein the sections are non-overlapping, vertical columns.

25. An interactive video system for allowing a user to select a view, the system comprising

a receiver for receiving encoded video content, the video content including a sequence of images, the images having been parsed into video streams representing sections of the images, each of the video streams having been individually compressed and formatted;

a decoder coupled to the receiver for extracting the individually compressed video streams; and

a processing unit coupled to the decoder for selectively decompressing the individually compressed video streams and merging the resulting image sections into seamless viewable video, the processing unit comprising user interface means for receiving user view selection instructions such that the processing unit selectively decompresses the individually compressed video streams to construct a view based on the user's view selection instructions.

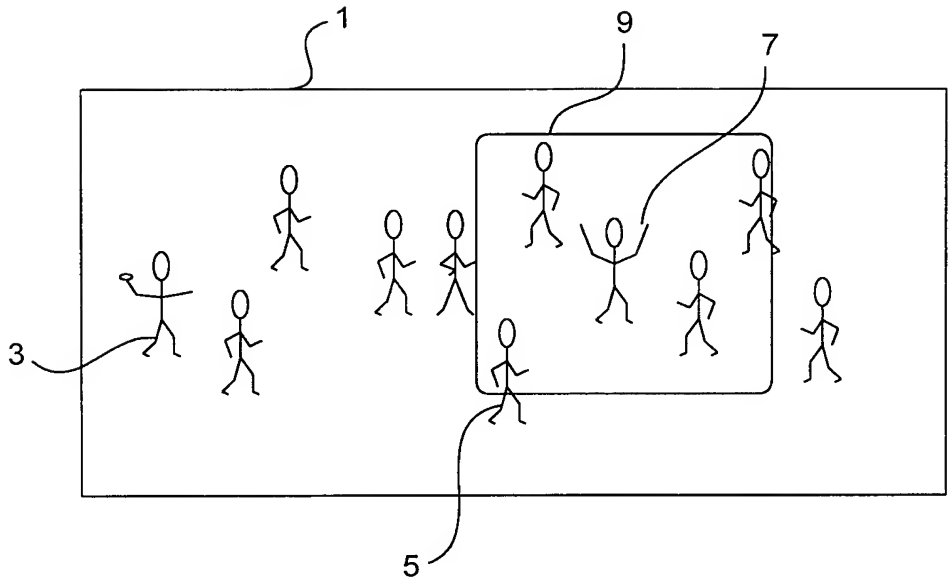


FIG. 1

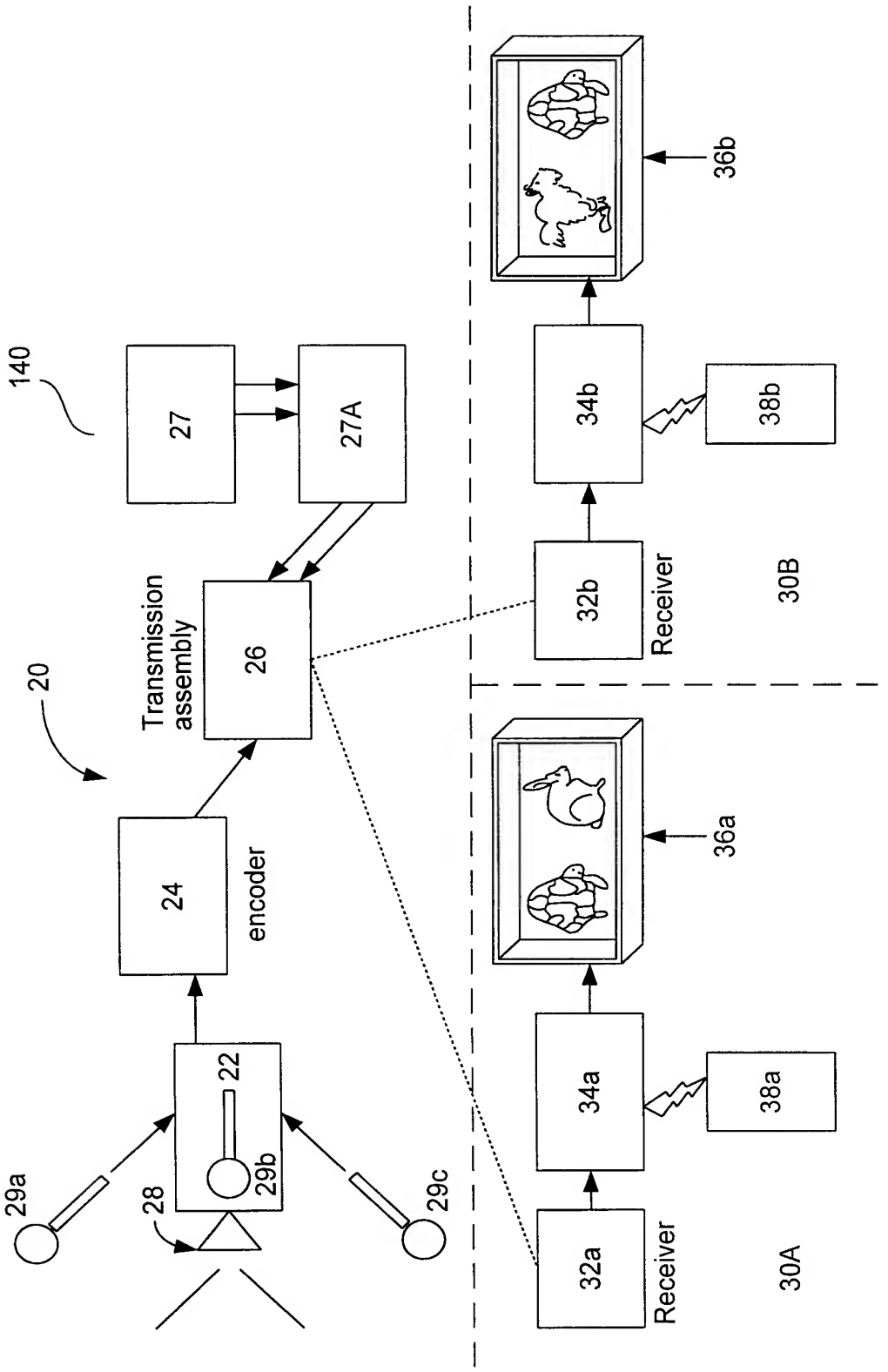

140
27
27A
Transmission assembly
26
20
24
encoder
22
29a
28
29b
29c
32a
Receiver
34a
38a
36a
32b
Receiver
34b
38b
36b
30A
30B

FIG. 2

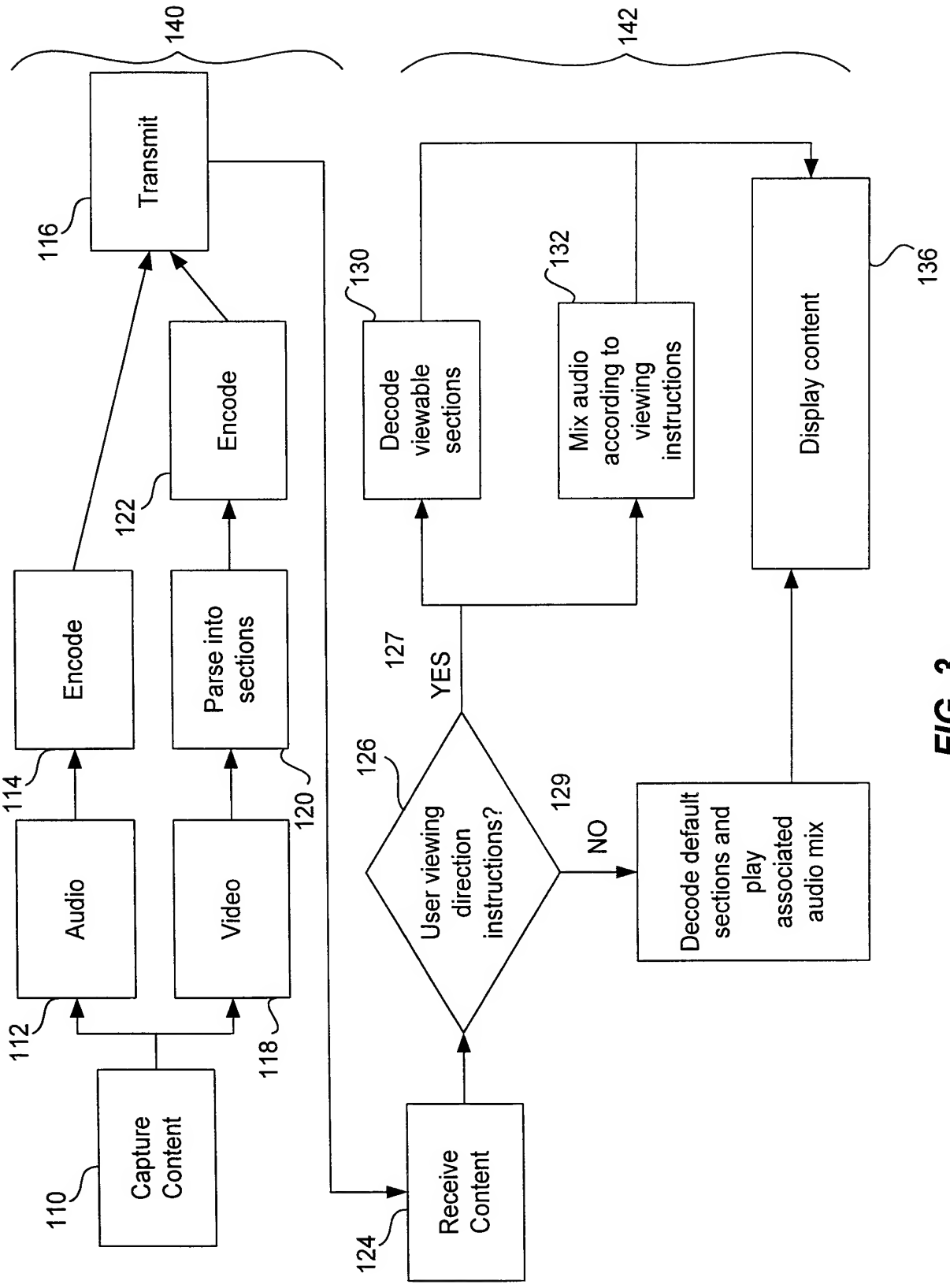

140
142
Transmit
116
Encode
122
Encode
114
Parse into sections
120
Audio
112
Video
118
Capture Content
110
Receive Content
124
User viewing direction instructions?
126
127
YES
129
NO
Decode viewable sections
130
Mix audio according to viewing instructions
132
Decode default sections and play associated audio mix
Display content
136

FIG. 3

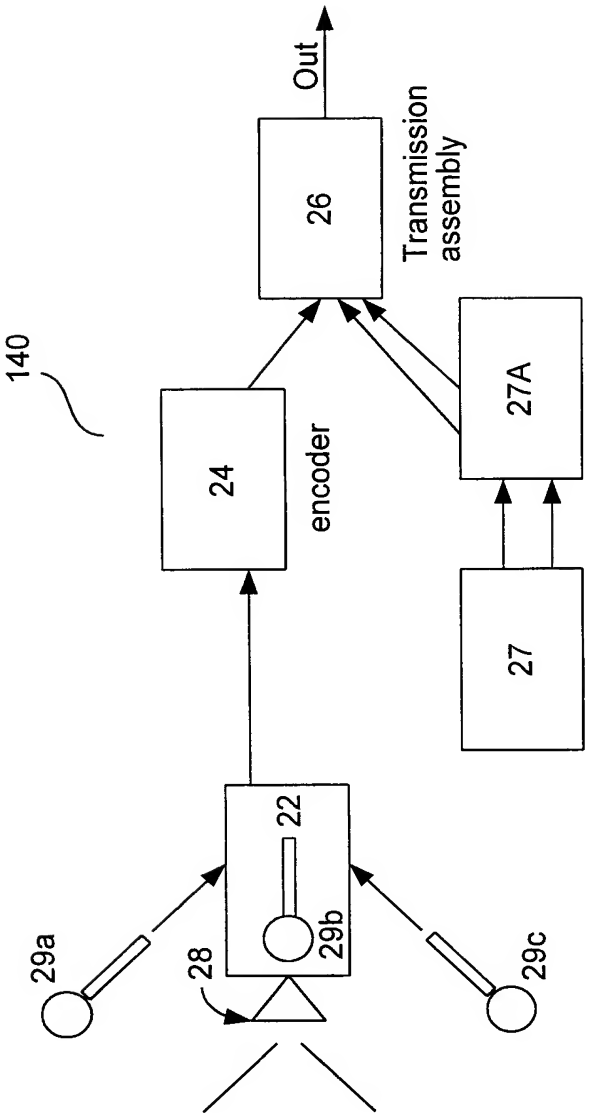

Out
26
Transmission assembly
140
24
encoder
27A
27
22
29b
29a
28
29c

FIG. 4

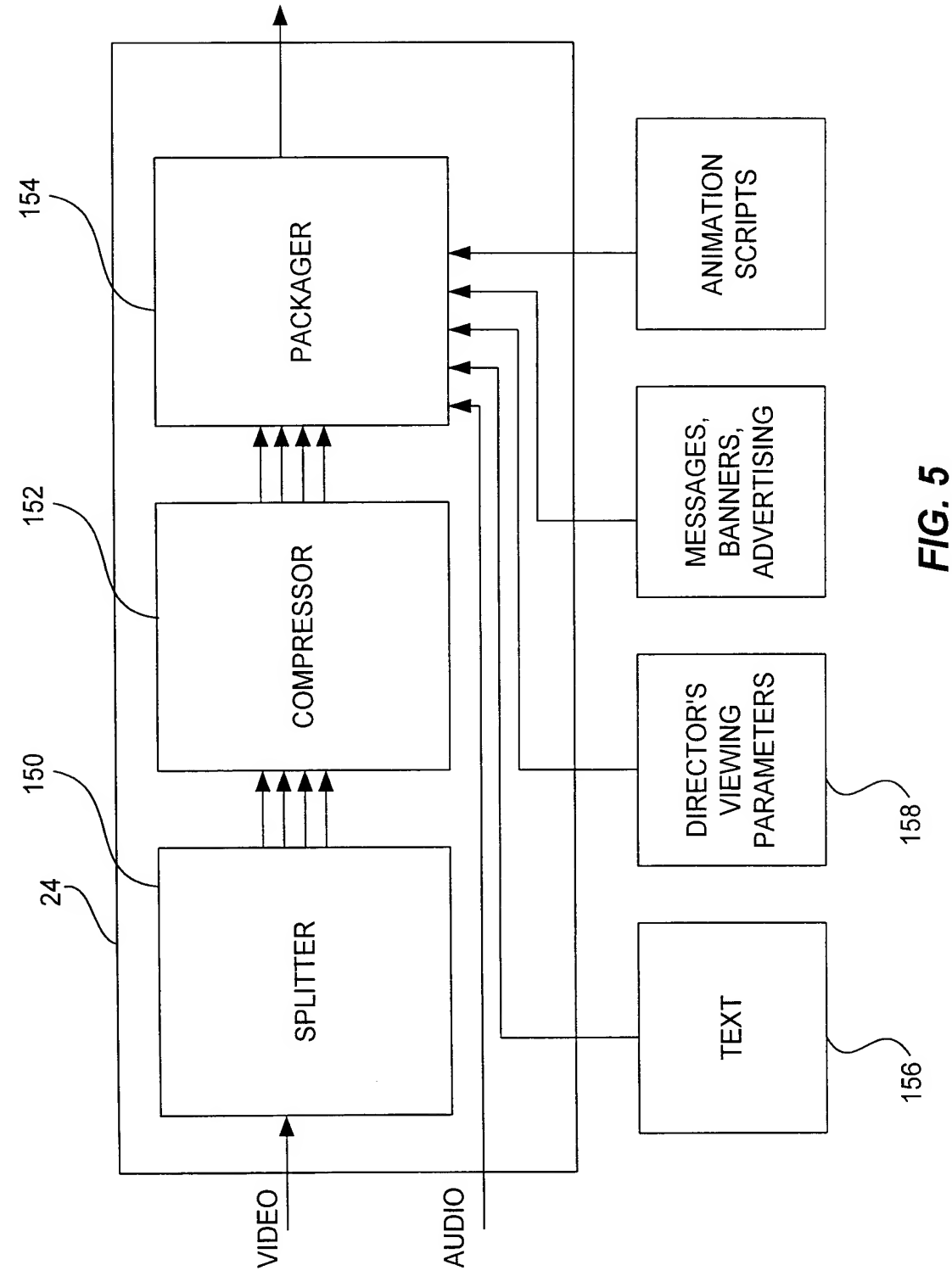

154
PACKAGER
152
COMPRESSOR
150
SPLITTER
24
VIDEO
AUDIO
ANIMATION SCRIPTS
MESSAGES, BANNERS, ADVERTISING
DIRECTOR'S VIEWING PARAMETERS
158
TEXT
156

FIG. 5

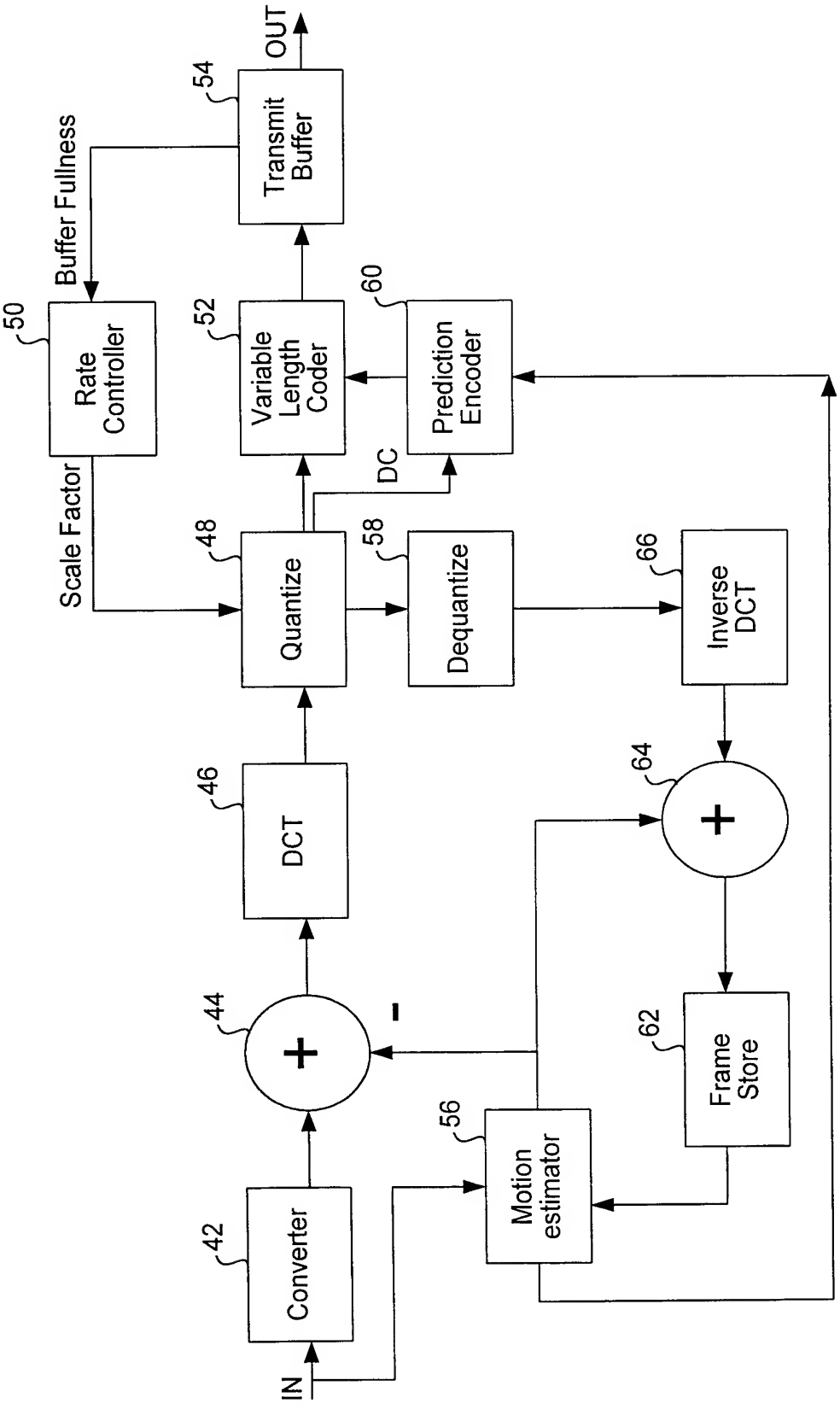

IN
Converter
42
DCT
46
Quantize
48
Dequantize
58
Inverse DCT
66
Variable Length Coder
52
Prediction Encoder
60
Transmit Buffer
54
OUT
Rate Controller
50
Buffer Fullness
Scale Factor
DC
Motion estimator
56
Frame Store
62
Motion Vectors
+
-
+
44
64

FIG. 6

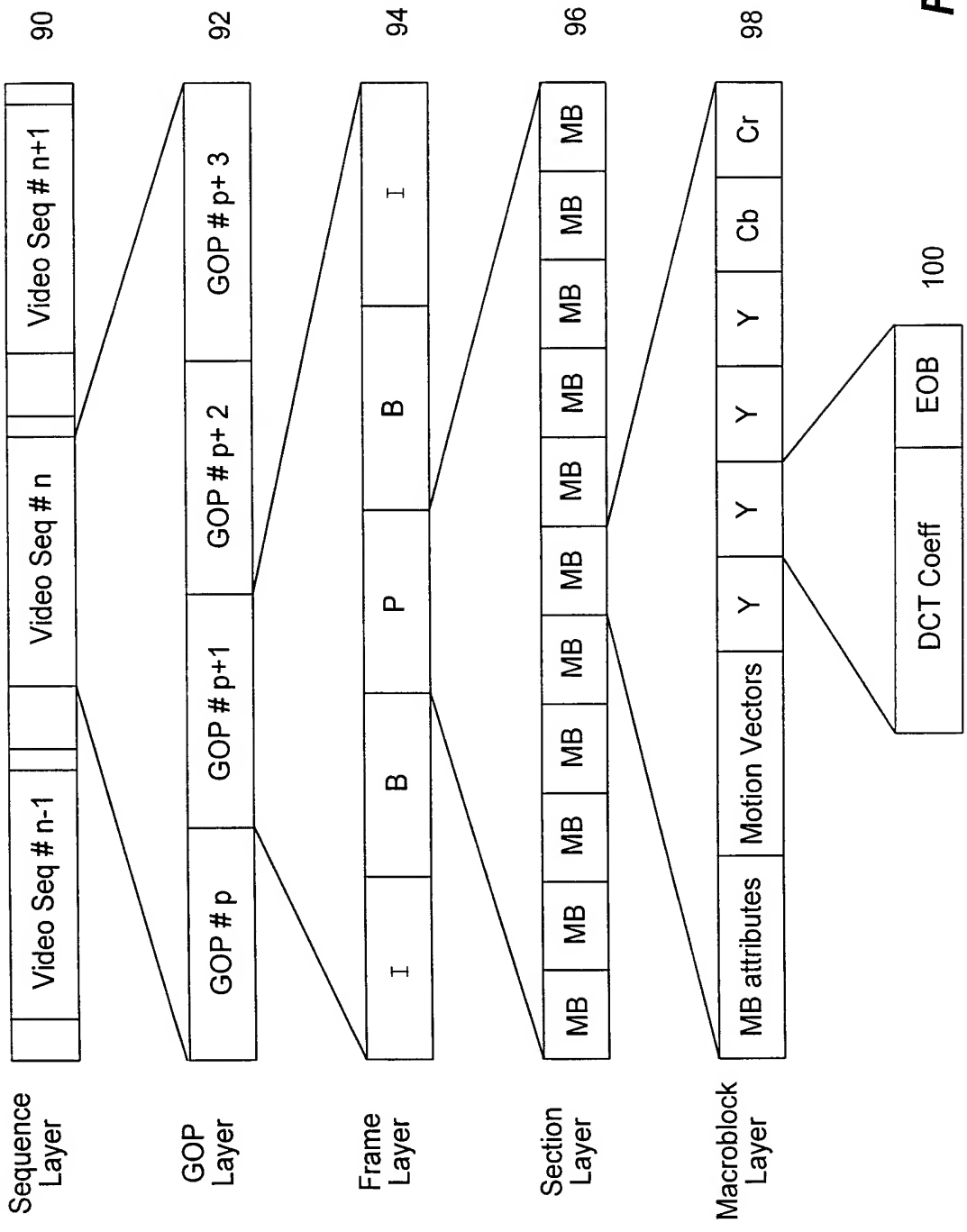

Sequence Layer
GOP Layer
Frame Layer
Section Layer
Macroblock Layer
90
92
94
96
98
100
Video Seq # n-1
Video Seq # n
Video Seq # n+1
GOP # p
GOP # p+1
GOP # p+ 2
GOP # p+ 3
I
B
P
B
I
MB
MB attributes
Motion Vectors
Y
Y
Y
Y
Cb
Cr
DCT Coeff
EOB

FIG. 7

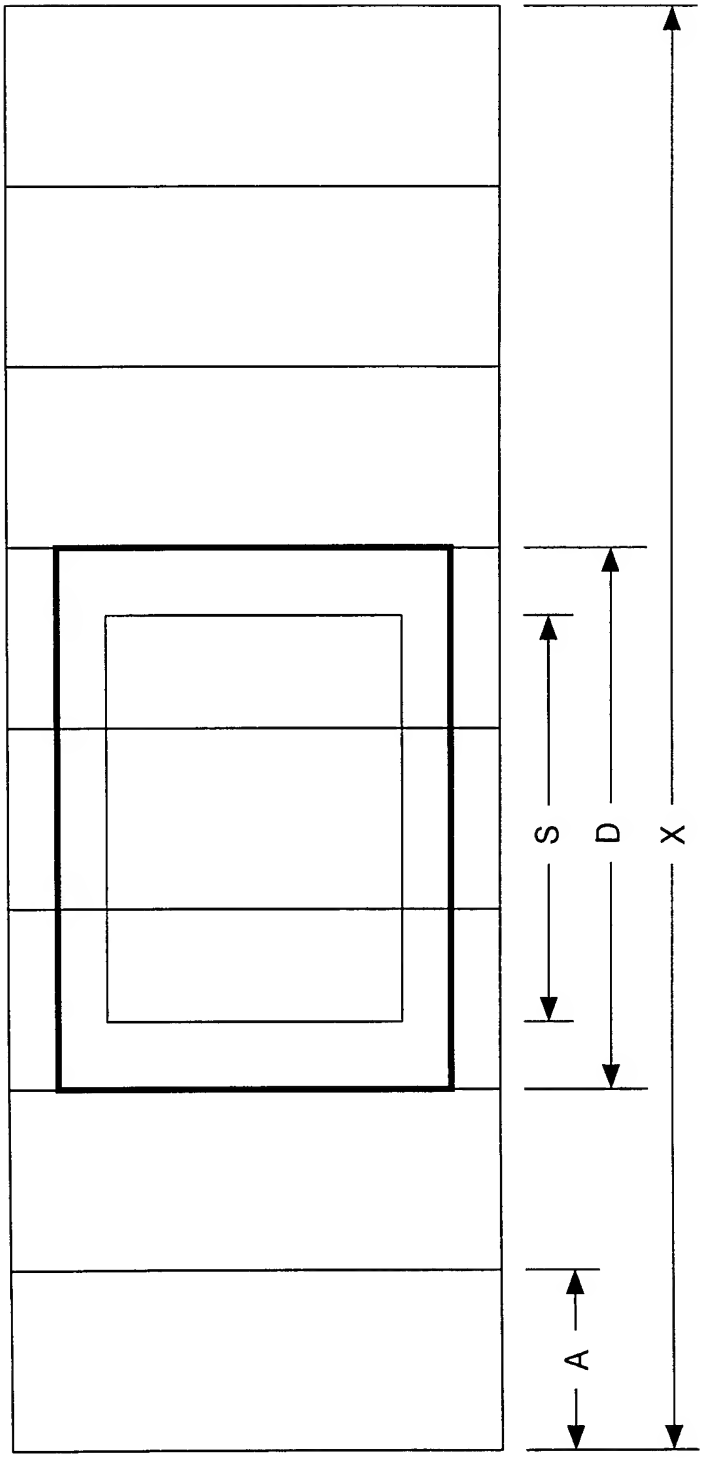


FIG. 8

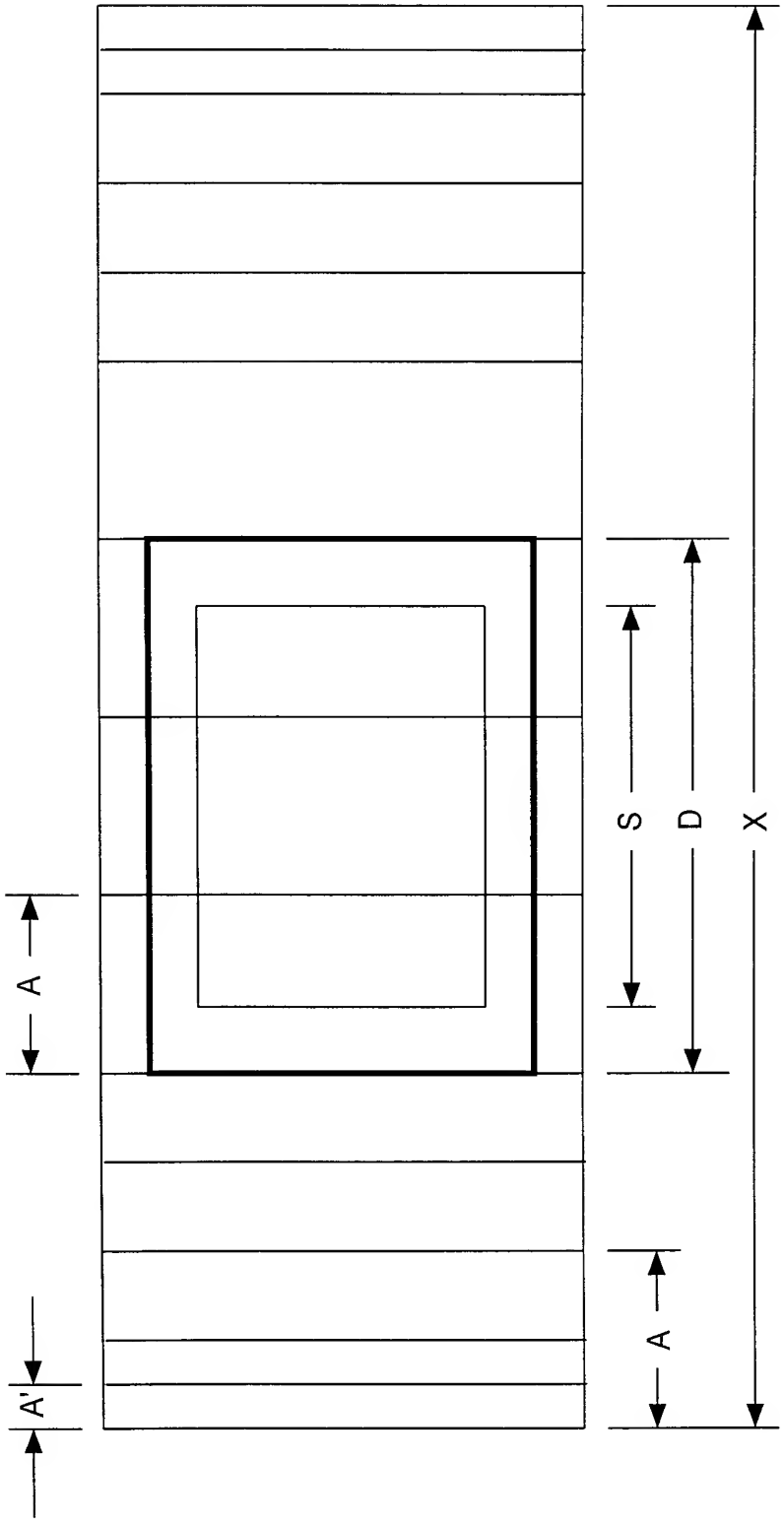


FIG. 9

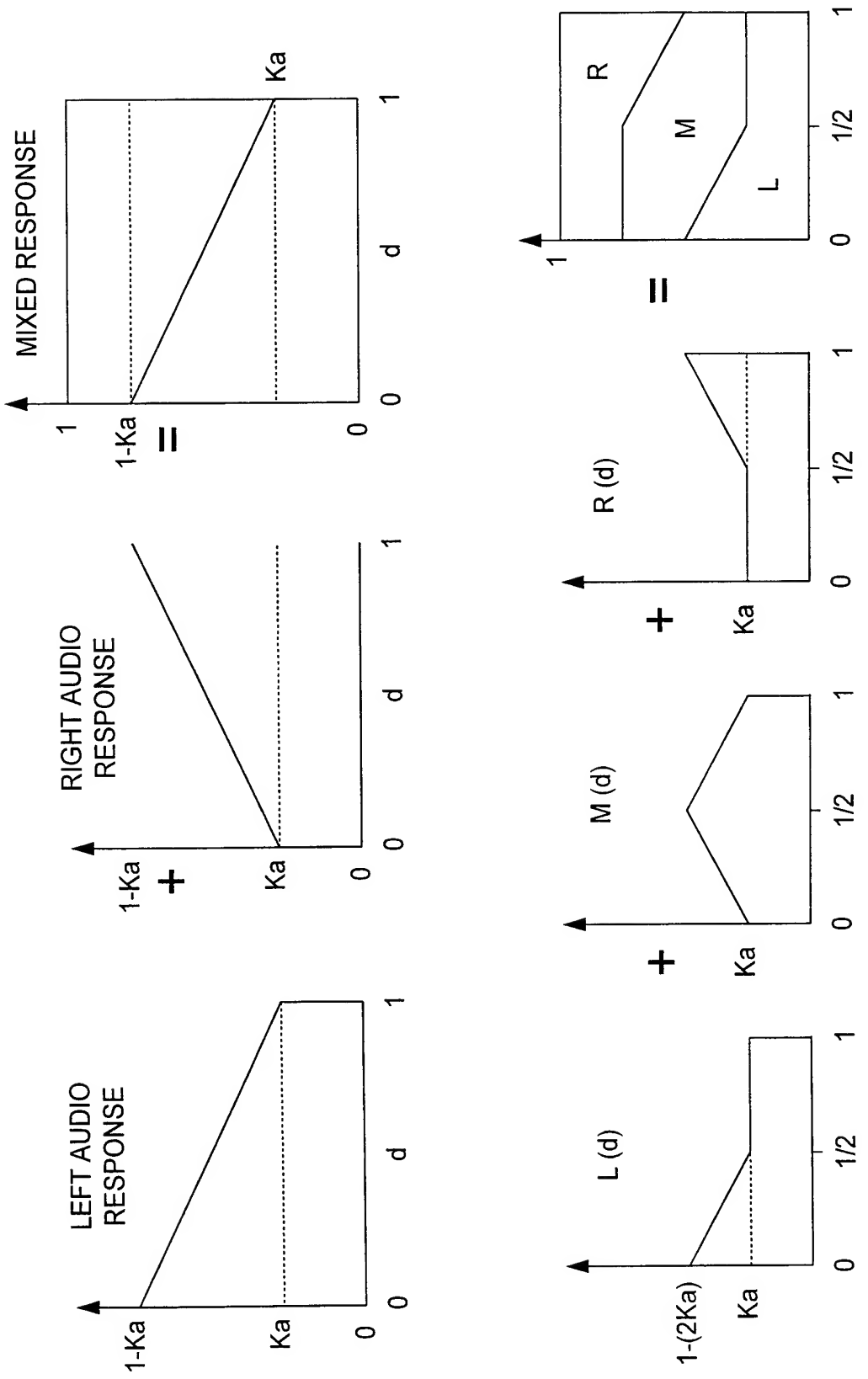

LEFT AUDIO RESPONSE
1-Ka
Ka
0
d
1
RIGHT AUDIO RESPONSE
1-Ka
+
Ka
MIXED RESPONSE
1
1-Ka
=
Ka
L (d)
1-(2Ka)
Ka
M (d)
+
Ka
R (d)
+
Ka
1
=
R
M
L
0
1/2
1

FIG. 10

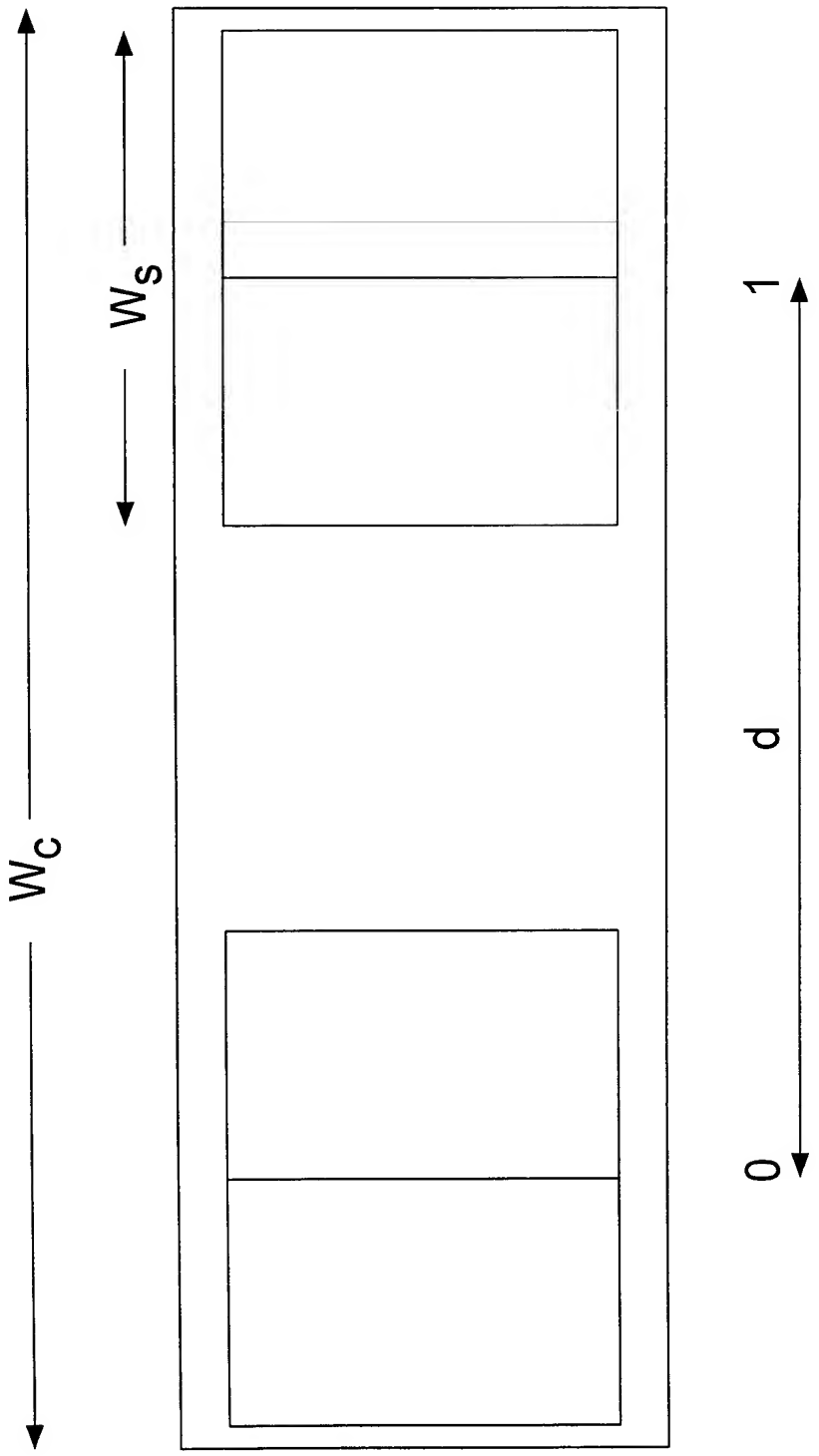


FIG. 11

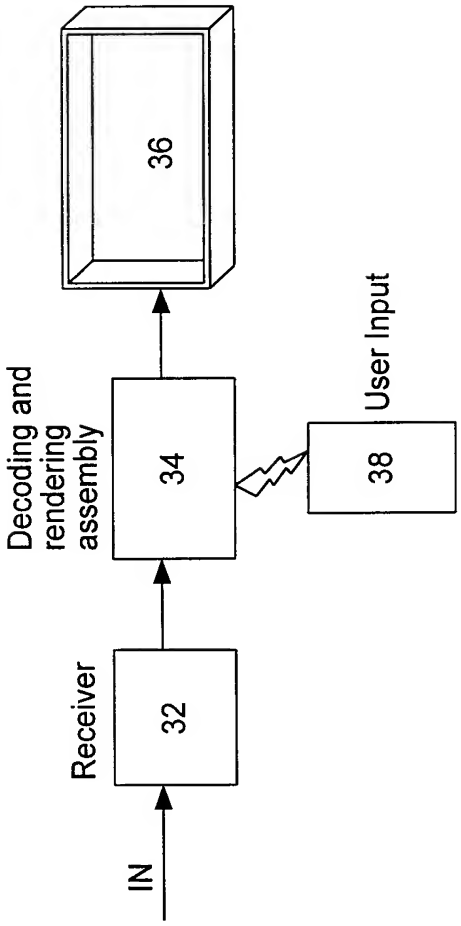

IN
Receiver
32
Decoding and rendering assembly
34
User Input
38
36

FIG. 12

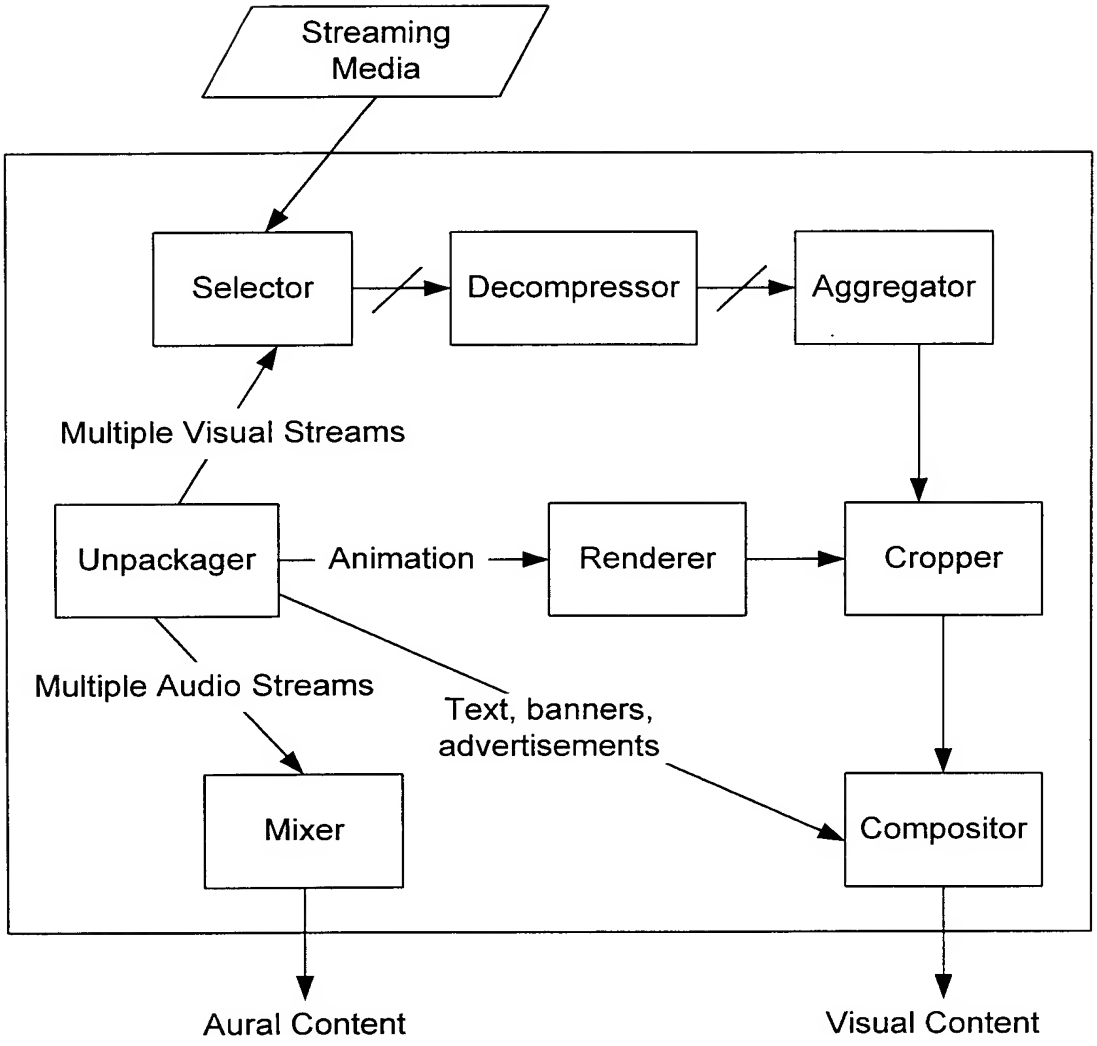

Streaming Media
Selector
Decompressor
Aggregator
Multiple Visual Streams
Unpackager
Animation
Renderer
Cropper
Multiple Audio Streams
Text, banners, advertisements
Mixer
Compositor
Aural Content
Visual Content

FIG. 13

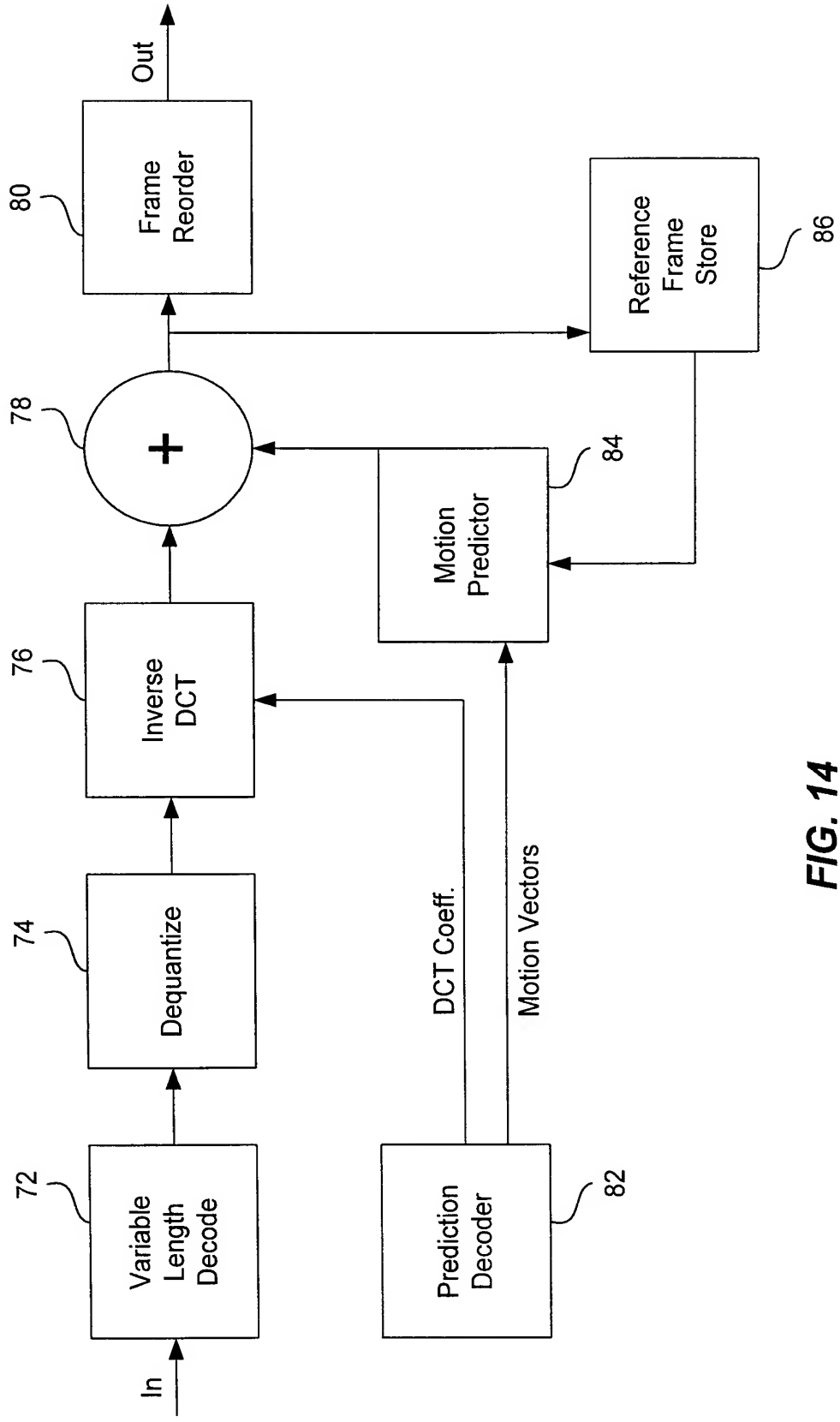

In
Variable Length Decode
72
Dequantize
74
Inverse DCT
76
+
78
Frame Reorder
80
Out
Prediction Decoder
82
Motion Predictor
84
Reference Frame Store
86
DCT Coeff.
Motion Vectors

FIG. 14

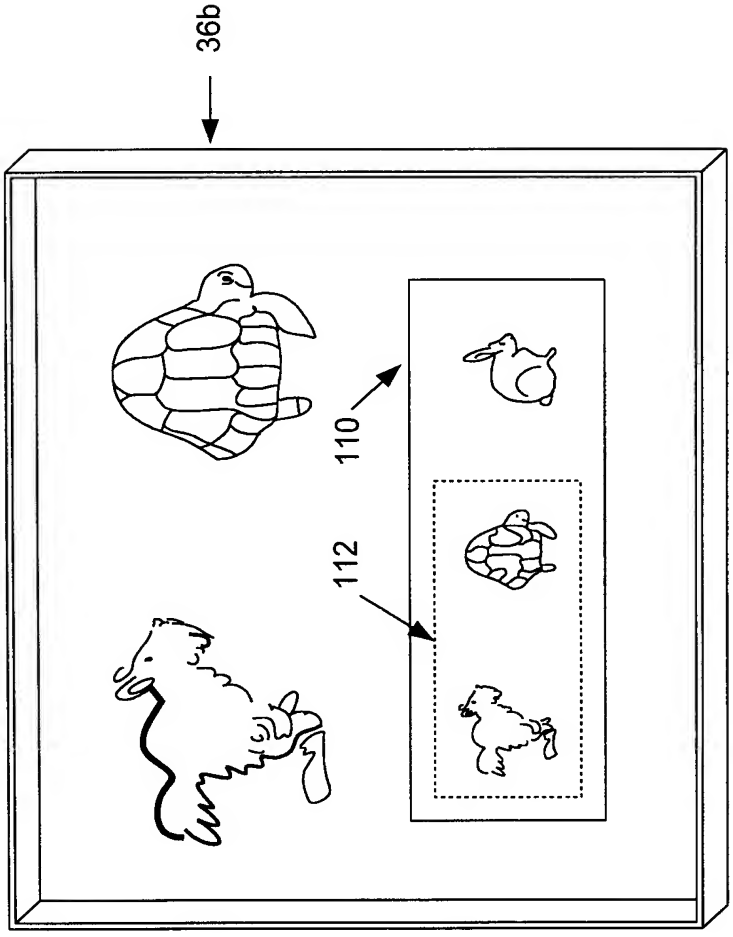


FIG. 15

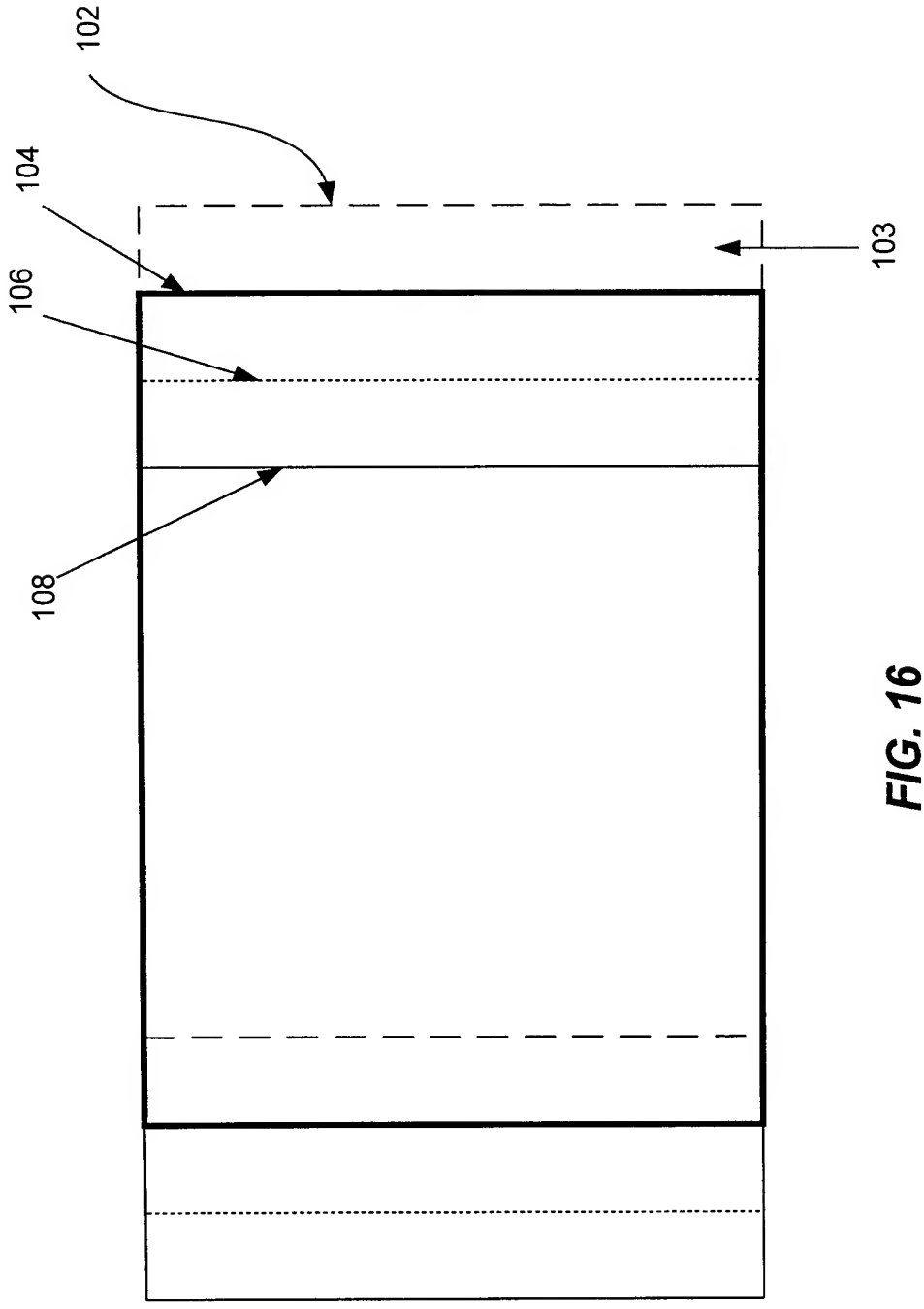


FIG. 16

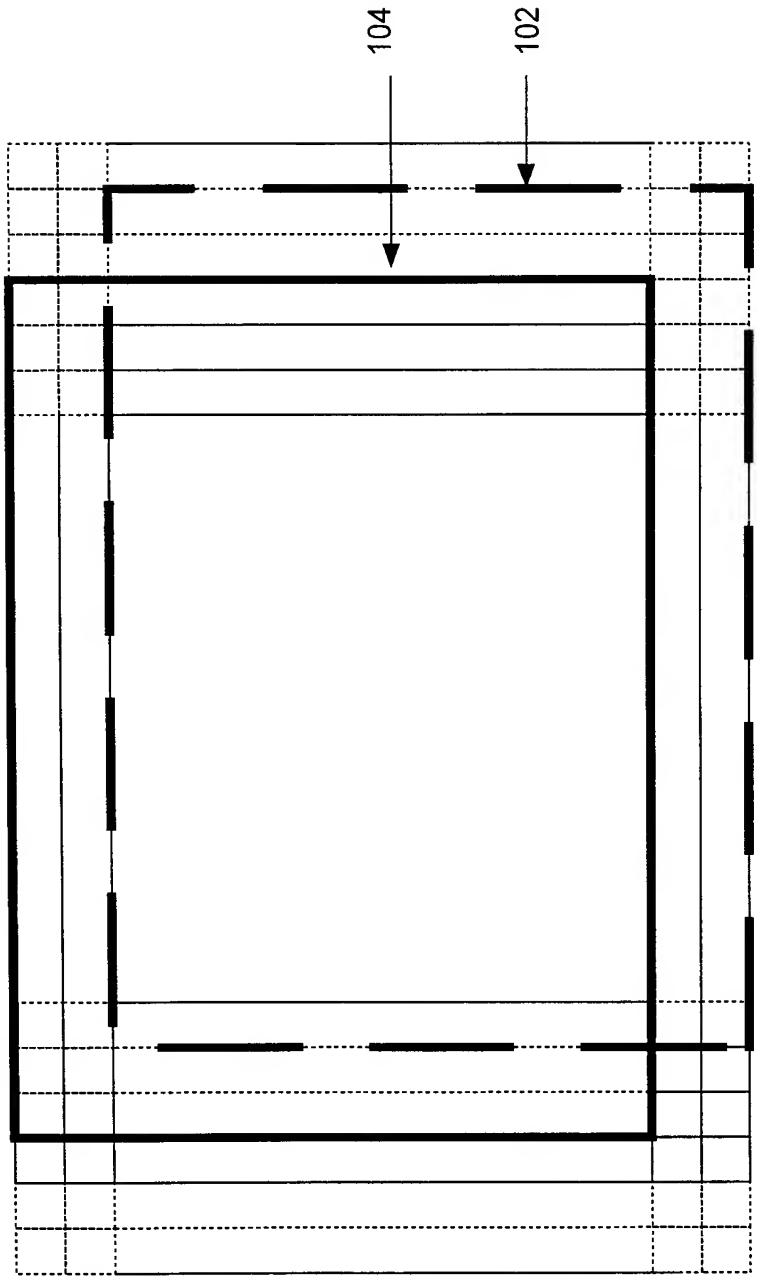


FIG. 17

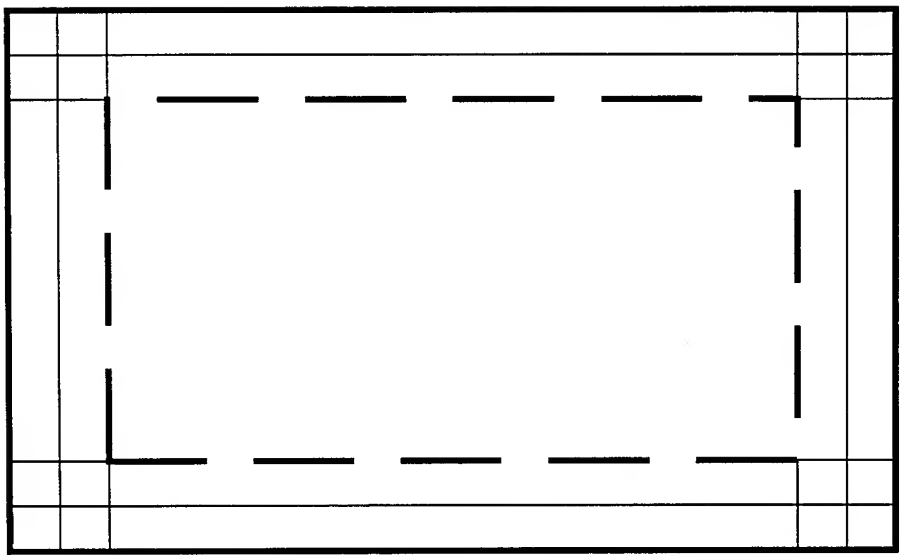


FIG. 18

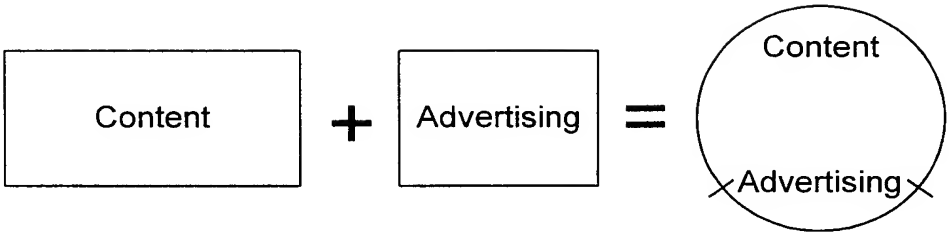


FIG. 19a

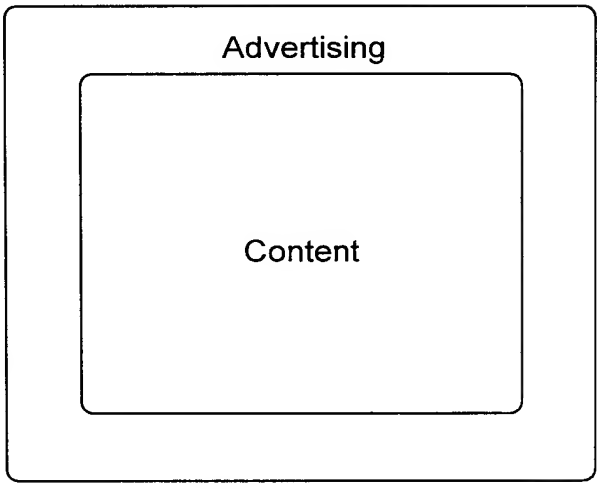


FIG. 19b